

Evaluating the measuring properties of the Principal Instructional Management Rating Scale in the Chinese Educational System: Implications for measuring school leadership

Panayiotis Antoniou¹ & Mohan Lu

University of Cambridge, Faculty of Education

Abstract

During the last 25 years researchers proposed a number of conceptual frameworks to measure the various functions of instructional leadership. One of the most frequently used frameworks is the Principal Instructional Management Rating Scale (PIMRS). Despite the great number of studies employing the PIMRS, evidence for its reliability and validity is relatively limited. In addition, we still don't know much in relation to the extent to which this instrument could be used in diverse demographic and cultural educational settings. This study explores the content, face, construct validity and reliability of the PIMRS in the Chinese Educational System. A total number of 311 teachers from five middle schools in Haidian District of Beijing participated in the study. The data were analysed using Confirmatory Factor Analysis. On an overall basis the results provided support to the face, content and construct validity and internal consistency of the PIMRS. However, six out of the fifty items had to be removed to reach satisfactory fit indices. Implications of the findings in relation to the importance of evaluating the measuring properties of research instruments are discussed and suggestions for future studies are finally provided.

Key Words: Instructional Leadership, PIMRS, school leadership in China, evaluation of measuring properties, Confirmatory Factor Analysis.

1. Introduction

Instructional leadership is identified in the literature as a factor associated with effective schools, by improving quality of teaching, which in turn enhance student

¹ Corresponding author's postal address: Faculty of Education, University of Cambridge, 184 Hills Road, CB2 8PQ, Cambridge, England, Tel: +44(0)1223 767554, Email:pa241@cam.ac.uk

learning (Bush, 2015; Antoniou, 2013b). Instructional leadership has been defined in a number of different ways, some of which refer to activities directly and others indirectly related to the processes of teaching and learning (Shatzer et al., 2014; Marks & Printy, 2003). According to Portin et al. (2003), Instructional leadership is the process of “assuring quality of instruction, modeling teaching practice, supervising curriculum, and assuring quality of teaching resources” (p. 18). It involves a number of functions such as coaching, critical reflection, teacher collaboration, teachers as action researchers and generally collaborative and critical thinking on quality of teaching (Glanz & Neville, 1997).

Many studies have been conducted during the last years to enhance our understanding and importance of instructional leadership (Hallinger & Lee, 2014; Bush & Glover, 2014). According to Leithwood et al. (2004), instructional leadership was seen as having an indirect impact on student outcomes through improving organizational learning culture and staff performance. Likewise, Marks and Printy (2003) found that school effectiveness could be improved by adopting instructional leadership. Robinson et al. (2008) focused on another point, concluding that transformational leadership associated with instructional leadership could have a significant influence on student outcomes and achievements. Lee et al. (2012) have summarized the two main conclusions that stem from all previous studies on the concept. The first was that instructional leadership affected learning organizations’ performances in a positive way. The second was that instructional leadership is a multifaceted structure that adjusts differently depending on the context, which is of great importance to this study.

During the last 25 years researchers proposed a number of conceptual frameworks aiming to describe the various functions of instructional leadership (e.g., Bossert, Dwyer, Rowan, & Lee, 1982; Hallinger & Murphy, 1985; Pitner, 1988; Leithwood & Stager, 1989; Ogawa & Bossert, 1995). In addition to the conceptual progress, methodological progress has also been made through the development of improved data collection instruments measuring instructional leadership (e.g., Leithwood & Steinbach, 1991). One of the most renowned and frequently used conceptual frameworks of instructional leadership is the Principal Instructional Management Rating Scale (PIMRS) (Hallinger, 1983; Hallinger & Murphy, 1985). During the last years, the PIMRS has been used extensively in more than twenty-five countries and in more than two-hundred empirical studies (Hallinger & Wang, 2013). According to Hallinger et al., (2013), the PIMRS has maintained a consistent record of yielding reliable and valid data. In particular and in relation to the PIMRS reliability, Hallinger (2008) states that “while relatively few researchers using the instrument sought to replicate the initial findings, several did. The replication studies of reliability and validity included Howe (1995), Jones (1987), Nogay (1995), Sawyer (1997), Taraseina (1993), and Wotany (1999)” (p. 24).

However, the earlier review by Hallinger (2011b) examined broadly the various methodologies that have been employed, utilising the PIMRS. As a result, very few information was provided in relation to the reliability and construct validity of the PIMRS based on the results obtained in previous studies. A number of researchers (e.g., Krug, 1990) have discussed some of the main issues related with the quality of the PIMRS. For example, a criticism relates to the length and complexity of the items. The main criticism is related to the lack of supportive evidence in relation to its reliability and validity which are sometimes hard to assess mainly because of

insufficient size of research samples. Most importantly, we still don't know much in relation to the extent to which the PIRMS is sensitive and appropriate to use in schools of different and diverse student population, schools of different size, of different levels of education (from primary to secondary) etc. In addition, it is still not clear whether the instrument is sensitive to diverse contextual factors which could modify the interpretation of the items included in the PIMRS (Condon & Matthews 2010). The above issues stress the importance of an updated evaluation of the validity and reliability of the PIMRS through appropriate statistical approaches such as the Confirmatory Factor Analysis.

In addition, very little information is available in relation to the measuring properties, and especially about the construct validity, of the PIRMS in the Chinese Educational System. As Hallinger et al., (2013) acknowledged, from the rather limited number of studies evaluating the validity of the PIMRS, three were conducted in the United States (Hallinger, 1983; Howe, 1995; Jones, 1987), one in Thailand (Taraseina, 1993), and one in Cameroon (Wotany, 1999). The overall conclusion from the literature is that there is a general shortage of research on leadership in the Chinese context (Walker & Dimmock, 2002). Despite the growing interest and writings on school leadership in the Chinese context (Sun, 2014), most studies have explored the links between headteacher's leadership and student outcomes in a western context, and school leadership has been mainly constructed and developed theoretically by Western scholars (Bush, 2014). Thus, we could claim that a greater focus on a cultural approach to exploring leadership must be undertaken because leadership in the West and in China possess different cultural roots (C.-C. Chen & Lee, 2008; Edwards & Turnbull, 2013). Leadership is a socially constructed process the essences of which are culturally affected. This is important as not only the conceptualisation of

leadership varies, but also the ways it is exercised varies across societal cultures (Dimmock & Walker, 2005; Yukl, 2006). As Dimmock (2011) argues “it is timely for educational researchers in Asia to generate cultural- and empirical-knowledge based in school leadership that will speak to the specific interests of Asian students, educators and practitioners” (p. 321).

2. Research Aims

Exploring the measuring properties of research tools is important and has important implications to the research design, the quality of the data and the conclusions drawn. Considering the importance attributed to measuring school leadership and identifying its relation to student outcomes internationally, a detailed and systematic evaluation of the measuring properties of widespread research instruments and tools is necessary for further improvements in this field. This is even more important in the case of the PIMRS, as no systematic attempt to evaluate its measuring properties has been undertaken since 1985 (Hallinger & Murphy, 1985).

In addition, the use of Confirmatory Factor Analysis approaches to evaluate the construct validity and internal consistency of the PIRMS has very rarely been reported. For those reasons, and also taking into consideration the contextual and cultural characteristics of the Chinese educational system, the purpose of this study is to explore the face, content and construct validity, reliability and internal consistency of the PIMRS research tool and provide suggestions to researchers, school headteachers and educators in relation to the extent to which the PIMRS could be used in the Chinese educational system to measure instructional leadership. In doing

so, the importance of considering contextual and cultural factors in choosing appropriate methodological tools to measure instructional leadership is also discussed.

3. Framework of the study - PIMRS

The Principal Instructional Management Rating Scale (PIMRS) was designed by Hallinger and Murphy (1985) and consists of three dimensions: (a) defining the School Mission, (b) managing the Instructional Program, and (c) developing a Positive School Learning Climate. Each dimension is further analysed into 10 instructional leadership functions. The respondents are requested to indicate the frequency of a headteacher's actions on a scale ranging from 1 (almost never) to 5 (almost always). Particularly, the "Defining the School Mission" dimension, consists of two functions, i.e., frames the school's goals and communicates the school's goals. Both of those functions are related with the extent to which a headteacher works with teachers to develop the school mission and the extent to which this mission is focused on student academic progress. The second dimension relates to the extent to which a headteacher coordinates the school instructional program. This incorporates three leadership functions: supervises and evaluates instruction, coordinates the curriculum, and monitors student progress. Finally, the third dimension of the PIMRS consists of several functions such as protecting teaching time, promoting teacher professional development, maintaining high visibility, providing incentives for teachers, and providing incentives for learning. This dimension has a broader focus and to a certain extent overlaps with factors related with transformational leadership (e.g., Leithwood et al., 2006; Marks & Printy, 2003).

4. Research Methods

Employing a survey research design, five middle (lower secondary) schools have been selected from the Haidian District of Beijing. From a total number of 544 teachers that were working at the five schools of our sample, we distributed 492 questionnaires. A total number of 311 questionnaires were returned (a response rate of 63.2%). Information related with the procedure employed to translate the PIMRS and the participants are provided below.

4.1 Translating the PIMRS

The questionnaire has been translated in Chinese following the approach proposed by Beaton et al., (2000). Particularly, for the forward translation step, we generated two translations of the original questionnaire by two independent translators, who are native speakers of the Chinese language. Following the comparison between the two independent translations, a reconciled language version has been developed along with a report elaborating on the reconciliation rationale. Then, for the backward translation step, the reconciled questionnaire in Chinese language was translated into English by one professional translator, native speaker of the Chinese language and fluent in English. The backward translation version and the original questionnaire have been finally compared. Some minor discrepancies that have been encountered have been resolved.

4.2 Sampling

Due to the size and complexity of the educational system in China, it was not possible to gain a representative sample of the whole country. For practical and accessibility

reasons, it was decided to focus on Haidian District of Beijing. Haidian is the second largest district in Beijing. It lies towards the northwestern part of the urban core and it is where most universities are located. Schools in Haidian district are considered as among the best schools and are usually found at the top of the education league in China.

The sampling approach was based on a random stratified approach, drawing from a list of the Haidian District records with all private schools allocated in three stratas / clusters i.e., (a) top performing schools, (b) average schools, and (c) poor performing schools according to their students' results in the National High School Entrance Examination. From a total number of 66 private schools, five schools were selected. The following table provides some basic information in relation to the characteristics of the schools in our sample.

.....

Insert Table 1 about here

.....

As we can observe from Table 1, two of the schools included in the sample were situated in the top 20 performing schools, two from the 20 poor performing schools, and the last one from the average schools cluster. The average “years of teaching” was 12.81, and the average “years of working with the current headteacher” was 4.1 years. These figures support that most of the participants were experienced teachers who were familiar with the school and their respective headteacher actions and behaviors,

and therefore able to make credible judgments in relation to the items included in the PIMRS.

5. Data Analysis

The data analysis was initially conducted for each one of the ten functions of the PIMRS. Then, the extent to which those functions could be incorporated into one of the three dimensions: (a) defining the School Mission, (b) managing the Instructional Program, and (c) developing a Positive School Learning Climate (Hallinger & Murphy, 1985) has been evaluated. In particular, the analysis evaluated the construct-validity, the reliability and the internal consistency of each function and dimension. The data were analysed using the SPSS (v.22), the SPSS-AMOS and the EQS software programs.

5.1. Content and Face Validity of the PIMRS

As the PIMRS has very rarely been used in the Chinese context, we considered important to explore the content and face validity of the questionnaire. The *content validity* was evaluated in collaboration with two faculty members of the Department of Education in a Chinese university and three experienced school headteachers, from schools not included in the research sample. This kind of validity refers to the extent to which the content of the measuring instrument is appropriate and relevant to the research purpose. Content validity indicates whether the content reflects the complete range of the attributes under study and is usually undertaken by a number of experts (Pilot & Hunger 1999; DeVon *et al.*, 2007; Antoniou, 2012). On an overall basis, the conceptual framework of the PIMRS was found to be comprehensive and satisfactory in relation to contemporary research findings on school leadership research and in

relation to leadership practices in Chinese middle schools. However, doubts have been raised in relation to the suitability of some items in the Chinese education system. Such items were mainly related with the flexibility of school headteachers to take decisions on issues prescribed by the MoE in China, such as the school curriculum and dealing directly with students and parents. Some suggestions relating to the extent to which some of the PIMRS items could be expanded so as to capture more accurately issues such as the student and teacher behaviour outside classrooms (during break-time) were provided, as part of school learning climate. At this stage, we decided to keep the PIMRS in its formal and complete version and explore those issues empirically via the results of the CFA analyses.

Secondly, the *face validity* of the surveys was examined. Face validity indicates the extent to which questionnaires appear to be suitable to a study's purpose and is considered as the weakest form of validity (Haladyna 1999; Trochim 2001; DeVon *et al.*, 2007). The three school headteachers, the two faculty members mentioned above and five teachers, who did not work to any of the sampled schools of the main data collection phase, were asked to evaluate the face validity of the PIMRS. All provided positive comments in terms of how the instrument appears on an overall basis, and in particular issues related with the questionnaire readability, style, language used and formatting.

5.2 Internal Consistency / Reliability of the PIMRS

To evaluate the *internal consistency* of the PIMRS, Cronbach's Alpha (1990) was calculated for each of the ten functions of the PIMRS. The results were particularly satisfactory results ($\alpha > 0.82$) for six out of the ten subscales. Further examination of "Cronbach's Alpha if Item Deleted" suggested that all items should be retained for

these six sub-scales. However, the “Cronbach’s Alpha if Item Deleted” for the remaining four functions, indicated that the reliability could reach particularly satisfactory results by deleting a number of items. Interestingly, almost all of those items have been identified during the evaluation of the PIMRS content validity as not particularly applicable for the Chinese context by our experts’ team. Based on the “Cronbach’s Alpha if Item Deleted” results, the six items, presented in Table 2 below, have been removed to reach satisfactory reliability indices.

.....

Insert Table 2 about here

.....

Once the six items have been taken out, the reliability for each function was recalculated. As demonstrated in Table 3, all functions were found to have α scores ranging from .82 to .94, which indicates high internal consistency in all cases.

.....

Insert Table 3 about here

.....

We also followed the suggestion by Hallinger et al., (2013) and employed the Ebel’s (1951) test to explore the reliability of the PIMRS. This test provides a reliability estimate which is based on the total - aggregated teacher responses from each school. Some researchers (e.g., Howe, 1995; Taraseina, 1993) suggested that Cronbach’s test violates a basic assumption by treating each teacher’s response independently, not

taking into account that teachers are grouped into schools. When employed the Ebel's formula to the data for each of the 10 leadership functions the reliability was again found to be satisfactory and higher than 0.84 for each of the ten functions.

5.3. Construct Validity of the PIMRS

To evaluate the *construct-validity* of the PIMRS, data were analysed through CFA approaches using the AMOS and EQS software programs. For each one of the ten leadership functions of the PIMRS separate CFA analyses were conducted to help identify the extent to which the theoretical models described by the PIRMS came within acceptable fit indices and parameters. Here we drew on structural equation modeling (SEM). There are two major types of variables in SEM, observed (indicator) variables and latent (construct) variables. As Schumacker and Lomax (2004) argue, "latent variables are not directly observable and hence they are inferred constructs, based on the observed variables that were selected to define each latent variable" (p. 196). So, to operationalise the latent variables i.e., the ten functions of the PIMRS, the instrument items (predictors) were used. Missing values were less than 4% so the typical method of list-wise deletion was employed (Allison, 2002).

5.3.1. First-Order Factors: Construct validity of the PIMRS functions

Having prepared the database for the analyses, first-order CFA models were tested and compared to identify the final model with the optimum fit indices for each function included in the PIMRS. This procedure finally led to the development of ten first-order CFA models, one of each function, demonstrating the construct validity of

the questionnaire items in each function. An example relating to the “Frames the school’s goals” function is described below.

A first-order CFA model designed to test the multidimensionality of a theoretical construct (Byrne, 1998) was used. Particularly, the model aimed to evaluate the construct validity of the “frames the school’s goals” function of the “Defining the school’s mission” dimension. The model hypothesised that (a) the five variables (i.e., questionnaire items) could be explained by one factor and (b) each variable would have a non-zero loading on the factor that it was designed to measure, and zero loadings on other factors. The findings of the first order factor SEM analysis generally affirmed the theory upon which this function of the PIMRS was developed. Particularly, the scaled χ^2 for the one factor structure ($\chi^2 = 5.6$, $df = 2$, $p > .05$) did not reach statistical significance, the RMSEA was .012 and the CFI was .95, all meeting the criteria for acceptable level of fit. Figure 1 presents the one-factor model and the factor parameter estimates, all of which were statistically significant ($p < .001$).

.....

Insert Figure 1 about here

.....

A similar approach was used for each of the ten functions of the PIMRS. Particularly, two first-order CFA models were generated for the “defining the School Mission” function, three first-order CFA models were generated for the “managing the Instructional Program” function and five first-order CFA models were generated for the “developing a Positive School Learning Climate” function. To evaluate the construct validity of the PIMRS sub-domains, several fit indices were estimated for

each model, such as the significance of X^2 , the RMR, SRMR, GFI, CFI and RMSEA. In cases where the factor-loadings of questionnaire items were not found to be considerably high, alternative models which excluded the particular items were tested and compared with the original models. In all cases, however, it was found that the existing models yielded a better fit than alternative reduced models; thus, all questionnaire items were retained in the final CFA models.

We also considered important to compare the fit indices of the CFA models with and without the six items that have been previously excluded based on the Cronbach's alpha reliability test results. Such comparisons were made for four functions. Particularly, comparisons were made for the "coordinates the curriculum" function, with and without Q17 and Q20, for the "monitors student progress" function, with and without Q25, for the "maintains high visibility" function, with and without Q32, and for the "provides incentives for learning" function, with and without Q48 and Q49. In all cases, the results provided empirical support to the construct validity of the reduced scales of the PIRMS functions. Based on the Cronbach's alpha and first-order CFA modeling results, the six items have been removed from the second-order factor analyses elaborated below.

5.3.2. Second Order Factors: Construct validity of the PIMRS dimensions

The previous section presented the analysis results providing support to the construct validity of the ten functions of the PIRMRS. Factors obtained from survey correlations are called first-order factors, irrespective of whether they are orthogonal or oblique. Due to the high positive correlation coefficients between all first order factors, i.e., functions, of each dimension, ranging from 0.83 to 0.95 ($p < 0.05$), it was

decided to consider the development of second-order factors; i.e., factors that may be determined from the correlations of the first-order factors, in our case, the three dimensions of the PIRMS. Such factors are important for the interpretation of the correlated variables. Second-order CFA models are usually applied when a measurement instrument measures a number of constructs, related to each other and each of which is, in turn, measured by several items. The basic assumption is that these distinct but related constructs could be represented by one or more common higher order constructs (DeYoung *et al.*, 2002).

For example, for the purpose of this study, we tested whether there is a second order factor for the “Developing the School Learning Climate” dimension that underlies the five specific leadership functions, such as “protecting instruction time” and “providing incentives for teachers” (each assessed by multiple items) which are considered as lower of first order factors (Chen et al., 2005). The assumption is that the higher order factor could account for the commonality among the specific issues measured by the lower order factors.

Second-order CFA analyses were performed for each of the three PIMRS dimensions of instructional leadership. Figures 2, 3 and 4 below present the findings and the factor loadings for (a) defining the School Mission, (b) managing the Instructional Program, and (c) developing a Positive School Learning Climate respectively.

.....

Insert Figure 2 about here

.....

.....

Insert Figure 3 about here

.....

.....

Insert Figure 4 about here

.....

The following observations could be made from the above figures. Firstly, the standardized factor loadings were all positive (higher than 0.55). The only exception found was related with two items of the *Protects Instructional Time* function of the *Developing the School Learning Climate* dimension, with standardized factor loadings of 0.46. The standardized path coefficients between the first- and second- order factors were all higher than .83. All parameter estimates were statistically significant ($p < .001$). To test the fitting of each model presented above several fit indices were estimated, such as the significance of χ^2 , the RMR, SRMR, GFI, CFI and RMSEA. It was found that the existing models yielded a better fit than alternative reduced models. The results of the analyses provided empirical support for the construct validity of the PIMRS dimensions and the findings of the second order factor SEM analysis generally affirmed the theory upon which the PIMRS was developed. In all cases, although the scaled chi-square ($\chi^2=87.1$, $df=3$, $p<0.001$) was statistically significant, the values of RMSEA (ranging from 0.012 to 0.031) and CFI (ranging from 0.955 to 0.969) met the criteria for acceptable level of fit.

6. Discussion

This study has evaluated the face, content and construct-validity and reliability of the PIMRS in the Chinese educational system. Although the PIMRS has been utilised extensively in a number of countries (Hallinger, P. & Chen, J., 2015; Hallinger, 2011b; Hallinger & Wang, 2013), the instrument has scarcely been used for data collection in the Chinese educational system. This is important as exploring the measuring properties of research tools could have important implications to the research design, the quality of the data and the conclusions drawn (Fromm et al., 2016; Antoniou & Kyriakides, 2011;2013).

Data were collected from a number of 311 teachers from five middle schools in Haidian District of Beijing. The questionnaire has been translated in Chinese following the approach proposed by Beaton et al., (2000) with backward and forward blind translations. The content and the face validity of the questionnaire have been explored by a local Chinese team of scholars, school headteachers and teachers with satisfactory results. To evaluate the reliability of the PIRMS the Cronbach's alpha (1990) and the Ebel's reliability tests (1951) have been calculated. To evaluate the *construct-validity* of the PIMRS, data were analysed through Confirmatory Factor Analysis (CFA) approaches using the AMOS and EQS software programs. For each one of the ten leadership functions of the PIMRS separate CFA analyses were conducted to help identify the extent to which the theoretical models developed and used in the study came within acceptable fitting indices and parameters. Second-order factor CFA models were also tested for each of the three PIMRS dimensions, i.e.,

factors that may be determined from the correlations of the first-order factors (DeYoung *et al.*, 2002).

The use of Confirmatory Factor Analysis approaches to explore the construct validity of the PIRMS has very rarely been reported. This is important as several researchers questioned the reliability and validity of several leadership scales, including the PIMRS (e.g. Condon & Matthews, 2010). The CFA is particularly appropriate to evaluate the extent to which measures of a construct, such as instructional leadership in our case, are in line with our understanding of the nature of that construct (Kline, 2010). In this perspective, for the data analysis in this study not only first order CFA models, but also second order CFA models have been developed. As Dwyer and Oh (1987) suggest, a second-order Structural Equation model can be used to combine several related first-order latent variables (i.e., functions of the PIMRS) into a single higher-order latent variable (i.e., one of the three dimensions of the PIMRS) to simplify a structural equation model and a theoretical framework.

Based on the analyses findings, to improve the fitting of the model in the Chinese educational system, the PIMRS was reduced from 50 to 44 items. Particularly, it was found that two items under the *Coordinates the curriculum* function, i.e., (a) draw upon the results of school-wide testing when making curricular decisions and (b) participate actively in the review of curricular materials, one item under the *Monitor student progress* function, i.e., Inform students of school's academic progress, one item under the *Maintains High Visibility* function, i.e., Visit classrooms to discuss school issues with teachers and students and finally two items under the *Provides incentives for learning* function, i.e., (a) recognize superior student achievement or improvement by seeing in the office the students with their work and (b) contact

parents to communicate improved or exemplary student performance or contributions were not found to be relevant and appropriate for measuring instructional leadership in the Chinese context and had to be removed to improve both the reliability of the scales and the fitting of the CFA models.

It is also important to note that those items had also been identified by the Chinese local experts-team we collaborated with, during the content validity evaluation of the PIMRS. For example, in relation to the *Coordinates the curriculum* function they stressed that the two items mentioned above might not be relevant in the Chinese educational system because the curriculum is strictly set by the MOE and there is very little flexibility left to the school headteachers to implement modifications or improvements. Similarly, in relation to the *Monitor student progress* and *Provides incentives for learning* functions, the local experts' team explained that in China the "grade-centered" administration has shifted the authority and direct responsibility of supervising teachers and students to grade administrators, rather than to school headteachers directly.

The results of this study indicate that in our attempts to measure and evaluate the impact of instructional leadership we need to consider two types of leadership functions: Holistic and Contextual. The Holistic functions refer to those functions that are found to work in a number of educational systems, irrespective of the differences in context and the diverse cultural backgrounds (Kyriakides & Creemers, 2009; Antoniou et al., 2015). For example, in this study it was found that 44 out of the 50 items included in the PIMRS are relevant for measuring instructional leadership functions in the Chinese educational system, as found in other educational systems (Heck & Hallinger, 1998; Hallinger et al., 2013).

At the same time, there are also contextual functions to be taken into consideration in our attempts to measure instructional leadership (Crow, 2001). For example, in this study it was found that not all PIMRS items were appropriate for measuring instructional leadership in the Chinese Educational system, as we had to remove 6 (out of the 50 items) to reach acceptable and satisfactory fit indices in the CFA analyses. Such contextual factors are based on the assumption that no single set of administrative competences will be effective in all different schools and social contexts (Davis et al., 2005).

We need to acknowledge that school leadership is a dynamic and multidimensional concept, in which context has an important role. As Hallinger and Heck (1998) argue there is no universal paradigm or theory for examining organizational behavior that is valid in all social or organizational contexts. A similar argument has been made by colleagues of Indigenous research. Such kind of research requires location-specific contextual factors that must be indigenous, but the theoretical lens can be borrowed (Eacott & Asuga, 2014). Leadership is a socially constructed process the essences of which are culturally affected. This is important as not only the conceptualisation of leadership varies, but also the ways it is exercised varies across societal cultures (Bush, 2011; Dimmock & Walker, 2005; Yukl, 2006). Leaders that grow up in different societal cultures have different internalised values and beliefs and these values and beliefs tend to drive them to exercise leadership in particular ways. This is also in line with the results of numerous studies conducted in China which found important variations between leadership in Chinese and Western contexts (Law Wing-Wah, 2010).

Although a number of authors have argued about the need to conceptualise school leadership using more complex conceptualizations, studies exploring leadership and its impact are still focusing on certain activities in isolation from the complex context in which school leadership takes place (Yang, 2014). As Mulford (2008) argues, “successful leaders adapt and adopt their leadership practice to meet the changing needs of circumstances in which they find themselves. As schools develop and change, different leadership approaches will inevitably be required and different sources of leadership will be needed so that development work keeps moving” (p.48). Any single one-size-fits-all or adjectival approach to leadership will eventually limit, leadership effectiveness (Crawford, 2012).

In an attempt to understand better how instructional leadership is implemented and how such practices vary between different educational systems, contexts, and cultures and make an impact in the improvement efforts, we argue in this paper that school leadership must be conceptualized as a complex system rather than as a linear series of events or actions (Clarke & Collins, 2007; Collins & Clarke, 2008; Opfer & Pedder, 2011), consisting of both holistic and contextual functions.

In such complex and dynamic systems, the importance of context in our efforts to explore the impact of leadership needs to be acknowledged (Antoniou, 2013a). This is mainly because all leadership described as successful is contingent and research findings support that school leaders interact and behave in various manners, depending on the conditions they face at any given time, the teachers and stakeholders with whom they are interacting. As Gronn (2003) argues, research about the forms and effects of leadership is becoming increasingly sensitive to the contexts in which

leaders work and how, in order to be successful, leaders need to respond flexibly to their contexts. The results of this study are also in line with the claims by Li et al., (2012) who claim that given that almost all extant theories of management are built upon the philosophies and values of the West (Leung, 2012), indigenous research in non-Western cultures, which have different intellectual and cultural traditions, has immense potential to contribute to universal theories by modifying, enriching, or supplementing Western management theoretical concepts.

On an overall basis the results of this study provided support to the face, content and construct validity and internal consistency of the PIMRS in the Chinese Educational system. The fit indices of the CFA models with the 44 items demonstrated that the instrument holds together well as entity and as separate factors. The findings of the second order factor SEM analysis generally affirmed the theory upon which this function of the PIMRS was developed. At the same time, six items had to be removed, which indicates, as mentioned earlier, the important role of contextual and cultural factors in measuring instructional leadership. These results indicate the importance of evaluating the measuring properties such as the reliability, face, content and construct validity of any research instrument, especially when those are to be adopted and used in different settings which are not similar to the ones that the instruments has been developed. The results also stress the significance of considering contextual and cultural factors in future attempts to measure instructional leadership.

By no means could the sample of this study, drawn from one district only, represent the whole of China. Future studies could also build on the findings of this study to assess the external validity of the PIMRS by drawing on different samples and cultural contexts. In addition, the external validity of the PIMRS could also be

evaluated by comparing its results with the results from other instruments (Hallinger et al., 2013). In relation to the research design, we could argue that there is a need for mixed-methods research design projects, to measure both the generic and contextual factors of school leadership, since in the past both qualitative case studies and quantitative studies were conducted in isolation and in parallel routes. By combining both strategies in a mixed-methods designs future studies could overcome problems of external validity and generalizability, which is the case for case studies, and at the same time problems of interpretation and de-contextualization, which is the case for quantitative large-sample studies. Such studies could further enhance our understanding of the suitability and external validity of the PIMRS in measuring instructional leadership especially in diverse educational settings. This could also be useful to researchers and educators in choosing among research instruments to measure instructional leadership and in making methodological choices when using the PIMRS.

References

- Allison, P. (2002). Missing data. Thousand Oaks, CA: Sage.
- Antoniou, P. (2013a). A longitudinal study investigating relations between stages of effective teaching, teaching experience and teacher professional development approaches. *Journal of Classroom Interaction*, 48 (2), 25-40.
- Antoniou, P. (2013b). Development of Research on School Leadership Through Evidence-based and Theory-Driven Approaches: A Review of School Leadership Effects Revisited. *School Effectiveness and School Improvement*, 24 (1), 122-128.
- Antoniou, P. (2012). The Short- and Long- term Effects of Secondary Schools upon Students' Academic Success and Development. *Educational Research and Evaluation*, 18 (7), 621 -640.
- Antoniou, P., Kyriakides, L. & Creemers, B.P.M. (2015). A Dynamic Integrated Approach to Teacher Professional Development. Rationale and Main Characteristics. *Teacher Development*.19 (4), 535 - 552.
- Antoniou, P. & Kyriakides, L. (2013). A Dynamic Integrated Approach to Teacher Professional Development: Impact and Sustainability of the Effects on Improving Teacher Behavior and Student Outcomes. *Teaching and Teacher Education*, 29 (1), 1-12.
- Antoniou, P. & Kyriakides, L. (2011). The Impact of a Dynamic Approach to Professional Development on Teacher Instruction and Student Learning: Results from an Experimental Study. *School Effectiveness and School Improvement*, 22 (3), 291 – 311.
- Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine*. 2000; 25:3186 -91.
- Bossert, S., Dwyer, D., Rowan, B., & Lee, G. (1982). The instructional management role of the principal. *Educational Administration Quarterly*, 18(3), 34-64.
- Bush, T. (2015). Understanding instructional leadership. *Educational Management Administration Leadership*, 43(4), 487-489.
- Bush, T. (2014). Educational leadership and leadership development in Africa: Building the knowledge base. *Educational Management, Administration & Leadership*. 42(6), pp. 787–791.
- Bush, T. (2011). *Theories of Educational Leadership and Management*, Fourth Edition. London: Sage.
- Bush, T. & Glover, D. (2014). School leadership models: What do we know? *School Leadership and Management*, 34(5), 553–571.
- Byrne, B.M. (1998), Structural Equation Modeling with LISREL, PRELIS and SIMPLIS: Basic Concepts, Applications and Programming. Mahwah, New Jersey: Lawrence Erlbaum Associates.

- Condon, C., & Matthews, C. (2010). *Measuring principal performance: How rigorous are commonly used principal performance assessment instruments?* (A quality school leadership issue brief). Naperville, IL: Learning Point Associates.
- Chen, C.-C., & Lee, Y.-T. (2008). *Leadership and management in China – Philosophies, theories, and practices*. Cambridge, UK: University Press, Cambridge.
- Chen, F. F., Sousa, K. H., & West, S. G. (2005). Testing measurement invariance of second- order factor models. *Structural Equation Modeling*, 12(3), 471-492.
- Cheng, Y.C. (2002). Three waves of education reform: Internal, interface, and future. In Y.C. Cheng, M.M.C. Mok, & K.W. Chow (eds.) *Hong Kong and International educational policies: A handbook*. (Vol. 1, pp.3-30). Hong Kong: Hong Kong Institute of Education.
- Cheng, Y.C. & Walker, J. A. (2008). When reform hits reality: The bottle-neck effect in Hong Kong primary schools. *School Leadership and Management*. 28(5), 467-48.
- CPECC. (1985). *Decision on Educational System Reform*. From <http://www.moe.edu.cn/edoas/website18/info3318.htm>. Accessed 18/12/2006.
- Crawford, M. (2012). Solo and Distributed Leadership: Definitions and Dilemmas. *Educational Management, Administration & Leadership*, 40 (5), pp. 610–620.
- Cronbach, L. J. (1990). *Essentials of psychological testing* (3rd ed.). New York: Harper and Row.
- Crow, G. M. (2001), *School Leader Preparation: A Short Review of the Knowledge Base*, National College for School Leadership, available at www.ncsl.org.uk/researchpublications.
- Davis, S., L. Darling-Hammond, M. LaPointe and D. Meyerson (2005), *Review of Research, School Leadership Study: Developing Successful Principals*, Stanford, Educational Leadership Institute, commissioned by the Wallace Foundation, Palo, Alto, CA.
- DeVon, H. A., Block, M. E., Moyle-Wright, P., Ernst, D. M., Hayden, S. J., Lazzara, D. J. et al. (2007). A psychometric Toolbox for testing Validity and Reliability. *Journal of Nursing scholarship*, 39 (2), 155-164.
- DeYoung, C. G., Peterson, J. B., & Higgins, D. M. (2002). Higher-order factors of the Big Five predict conformity: Are there neuroses of health? *Personality and Individual Differences*, 33, 533-552.
- Dimmock, C. (2011) Formulating a research agenda in school leadership and organisational change for school improvement in Singapore. *School Leadership and Management*, 31(4), 321-338. (doi:10.1080/13632434.2011.606271)
- Dimmock. C. and Walker, A. (2005). *Educational Leadership: Culture and Diversity*. London: SAGE Publications.
- Dwyer, F. R. & Sejo, O. (1987). Output Sector Munificence Effects on the Internal Political Economy of Marketing Channels. *Journal of Marketing Research*, 24 (November), 347-358.

- Eacott, S. & Asuga, G.N. (2014). School leadership preparation and development in Africa: A critical insight. *Educational Management, Administration & Leadership*. 42(6), pp. 919–934.
- Ebel, R. (1951). Estimation of the reliability of ratings. *Psychometrika*, 16, 407-424.
- Edwards, G., & Turnbull, S. (2013). A cultural approach to evaluating leadership development. *Advances in Developing Human Resources*, 15(1), 46-60.
<http://doi.org/10.1177/1523422312467144>
- Fromm, G., Hallinger, P., Volante, P. & Wang, W.C. (2016). Validating a Spanish Version of the PIMRS: Application in National and Cross-national Research on Instructional Leadership, *Educational Management Administration and Leadership*, DOI: 1741143215617948
- Glanz, J. , & Neville, R. F. (1997). *Educational supervision: Perspectives, issues, and controversies*. Norwood, MA: Christopher-Gordon.
- Glickman, C.D. (1992). *Supervision in transition*. Alexandria, VA: ASCD
- Gronn, P. (2003) *The new work of educational leaders: Changing leadership practice in an era of school reform*. London: Paul Chapman.
- Haladyna, T. M. (1999). *Developing and validating multiple-choice test items*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Heck, R.H., & Hallinger, P. (2005). The Study of Educational Leadership and Management Where Does the Field Stand Today? *Educational Management Administration & Leadership* 33 (2), 229-244
- Hallinger, P. (2011a). Leadership for learning: Lessons from 40 years of empirical research. *Journal of Educational Administration*, 49(2), 125-142.
- Hallinger, P. (2011b). A review of three decades of doctoral studies using the Principal Instructional Management Rating Scale: A lens on methodological progress in educational leadership. *Educational Administration Quarterly*, 47(2), 271-306.
- Hallinger, P. (2008). *A review of PIMRS studies of principal instructional leadership: Assessment of progress over 25 years*. Paper prepared for presentation at the annual meeting of the American Educational Research Association (AERA) May 14, 2008, New York.
- Hallinger, P. (1983). *Assessing the instructional management behavior of principals* (Unpublished doctoral dissertation). Stanford University, Stanford, CA.
- Hallinger, P. & Chen, J. (2015). Review of research on educational leadership and management in Asia: A comparative analysis of research topics and methods, 1995–2012. *Educational Management Administration Leadership*, 43 (1), 5-27.

- Hallinger, P. & Lee, M. (2014). Mapping instructional leadership in Thailand: Has education reform impacted principal practice? *Educational Management Administration and Leadership*, 42(1), 6-29.
- Hallinger, P., Wang, W.C., & Chen C-W. (2013). Assessing the Measurement Properties of the Principal Instructional Management Rating Scale: A Meta- Analysis of Reliability Studies. *Educational Administration Quarterly*, 49(2) 272-309.
- Hallinger, P., & Wang, W.C. (2013). Measurement properties of the Principal Instructional Management Rating Scale. Technical Report, version 4.2. Hong Kong: Asia Pacific Centre for Leadership and Change, Hong Kong Institute of Education.
- Hallinger, P., & Murphy, J. (1985). Assessing the instructional leadership behavior of principals. *Elementary School Journal*, 86(2), 217-248.
- Howe, W. (1995). *Instructional leadership in catholic elementary schools: An analysis of personal, organizational, and environmental correlates* (Unpublished doctoral dissertation). Stanford University, Stanford, CA.
- Jones, P. (1987). *The relationship between principal behavior and student achievement in Canadian secondary schools* (Unpublished doctoral dissertation). Stanford University, Stanford, CA.
- Kline, R. B. (2010). *Principles and practice of structural equation modeling (3rd ed.)*. New York, New York: Guilford Press.
- Krug, S. E. (1990). *Leadership and learning: A measurement-based approach for analyzing school effectiveness and developing effective school leaders. Project Report (Report No. BBB28334)*. Urbana, IL: National Center for School Leadership. (ERIC Document Reproduction Service No. ED327950)
- Kyriakides, L. & Creemers, B.P.M. (2009). The effects of teacher factors on different outcomes: two studies testing the validity of the dynamic model. *Effective Education*, 1 (1), 61-85.
- Law, Wing-Wah. (2010). Culture and School Leadership in China: Exploring School Leaders' Views of Relationship- and Rule-based Governance. In Alexander Wiseman (ed.), *Educational Leadership: Global Contexts and International Comparisons*. Bingley: Emerald Publishing. pp. 303-341.
- Lee, M., Hallinger, P. and Walker, A. (2012), A distributed perspective on instructional leadership in International Baccalaureate (IB) schools, *Educational Administration Quarterly*, 48 (4), 664-698.
- Leithwood, K., Day, C., Sammons, P., Harris, A., & Hopkins, D. (2006). *Seven strong claims about successful school leadership*. Nottingham, UK: National College of School Leadership.
- Leithwood, K., Louis, K, Anderson, S. & Wahistrom, K. (2004). *Review of research: How leadership influence student learning?* Wallance Foundation.

Leithwood, K. and D. Duke. (1998). A Century's Quest To Understand School Leadership. Handbook of Research on Educational Administration. American Educational Research Association.

Leithwood, K., & Steinbach, R. (1991). Indicators of transformational leadership in everyday problem-solving of school administrators. *Journal of Personnel Evaluation in Education*, 4(3), 221-244.

Leithwood, K. A., & Stager, M. (1989). Expertise in principals' problem solving. *Educational Administration Quarterly*, 25(2), 126-161.

Leung, K. (2012). Indigenous Chinese Management Research: Like It or Not, We Need It. *Management and Organizational Review*, 8 (1), pp. 1-5.

Li, P.P., Leung, K., Chen, C.Chao, Luo, J.-D. (2012). Indigenous Research on Chinese Management: What and How. *Management and Organizational Review*, 8 (1), pp. 7-24.

Marks, H., & Printy, S. (2003). Principal leadership and school performance: An integration of transformational and instructional leadership. *Educational Administration Quarterly*, 39(3), 370-397

MoE. (2006). *National Education Development Bulletin 2005*. From <http://www.moe.edu.cn/edoas/website18/info20464.htm>. Accessed 06/03/2007.

MoE. (2004). *Education Revitalization Plan 2003-2007*. From <http://www.moe.edu.cn/edoas/website18/info3338.htm>. Accessed 09/09/2006.

MoE. (2002). *Views on Reinforcing the Improving the Training of School Principals*. From <http://www.moe.edu.cn/jsdwiwu/gbpeixun/00.htm>. Accessed 10/09/2006.

MoE. (2001). *National Curriculum for Qualification Training of School Principals*. Beijing: Ministry of Education.

Mulford, B. (2008). *The Leadership Challenge: Improving learning in schools*. Australian Council for Educational Research, Victoria, Australia.

Nogay, K. (1995). *The relationship of superordinate and subordinate gender to the perceptions of leadership behaviors of female secondary principals* (Unpublished doctoral dissertation). Youngstown State University, Youngstown, OH.

Ogawa, R., & Bossert, S. (1995). Leadership as an organizational quality. *Educational Administration Quarterly*, 31(2), 224-243.

Pearl, Judea (2000). *Causality: Models, Reasoning, and Inference*. Cambridge University Press. ISBN 9780521773621.

Pilot, D., & Hunger, B. (1999). *Nursing research: principals and methods*. Philadelphia: Lippincott Williams & Wilkins.

Pitner, N. (1988). The study of administrator effects and effectiveness. In N. Boyan (Ed.), *Handbook of research in educational administration* (pp. 106-132). New York, NY: Longman.

- Portin, B., DeArmond, M., Gundlach, L., & Schneider, P. (2003). A Practical Look at School Leadership, *Policy Brief*, Center on Reinventing Public Education.
- Reitzug, U.C. & Cross, B. (1993). Deconstructing principal instructional leadership: from 'super' vision to critical collaboration, paper presented at the *annual conference of the University Council for Educational Administration*, Houston, TX.
- Robinson, V. M. J., Lloyd, C., & Rowe, K. J. (2008). The impact of leadership on student outcomes: An analysis of the differential effects of leadership type. *Educational Administration Quarterly*, 44(5), 635–674.
- Sawyer, E. (1997). *The relationship between cognitive styles and instructional leadership*. Unpublished doctoral dissertation. University of Missouri, Columbia, MO.
- Schumacker, R.E., & Lomax, R.G. (2004). *A Beginner's Guide to Structural Equation Modelling* (2nd ed). Lawrence Erlbaum Associates Inc., Mahwah, NJ.
- Shatzer, H. R., Caldarella, P., Hallam, R. P. & Brown, L.B (2014). Comparing the effects of instructional and transformational leadership on student achievement: Implications for practice. *Educational Management Administration and Leadership*, 42(4), 445-459.
- Sun, M. (2014). 30 years of Chinese educational management: achievements, characteristics and problems. *International Journal of Educational Management*, 28 (3), pp. 340-348.
- Taraseina, P. (1993). *Assessing instructional leadership behavior of secondary school principals in Thailand* (Unpublished doctoral dissertation). Peabody College for Teachers of Vanderbilt University, Nashville, TN.
- Trochim, W.M.K. (2001). *The Research Methods Knowledge Base*. Cincinnati: Atomic Dog.
- Walker, A. and Dimmock, C. (2002). Cross-Cultural and Comparative Insights into Educational Administration and Leadership: An Initial Framework. In Allan Walker and Clive Dimmock (eds). *School Leadership and Administration: Adopting a Cultural Perspective*. New York: RoutledgeFalmer. (pp.13-32).
- Wotany, J. (1999). *The training, selection, and hiring practices of Cameroonian school leaders and the evaluation of the present effectiveness of their instructional leadership behaviors* (Unpublished doctoral dissertation). University of Missouri, Kansas City, MO.
- Yang, Y. (2014). Principals' transformational leadership in school improvement. *International Journal of Educational Management*, 28 (3), pp. 279-288.
- Yukl, G. (2006). *Leadership in organizations, Sixth edition*. Upper Saddle River, NJ: Prentice Hall.